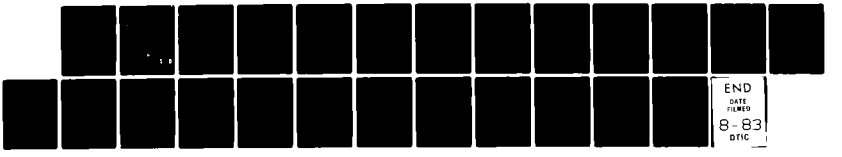
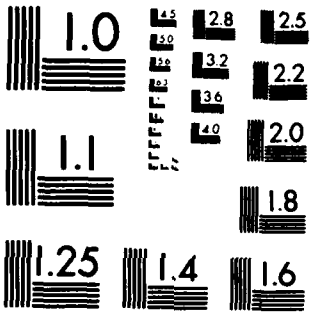
AD-A129 825 2400- TO 800-B/S LPC (LINEAR PREDICTIVE CODER) RATE CONVERTER(U) NAVAL RESEARCH LAB WASHINGTON DC L J FRANSEN 16 JUN 83 NRL-8716 1/1

UNCLASSIFIED F/G 9/5 NL



END
DATE FILMED
8-83
DTIC



NRL Report 8716

2400- to 800-b/s LPC Rate Converter

L. J. FRANSEN

Communications Systems Engineering Branch
Information Technology Division

June 16, 1983

NAVAL RESEARCH LABORATORY
Washington, D.C.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Report 8716	2. GOVT ACCESSION NO. AD-A129 825	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 2400- TO 800-b/s LPC RATE CONVERTER	5. TYPE OF REPORT & PERIOD COVERED Final report on a continuing NRL problem.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) L. J. Fransen	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, DC 20375	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 33904N X7290-CC 75-0114-0-3
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE June 16, 1983
	13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.	
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)	
18. SUPPLEMENTARY NOTES	
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Speech encoder Vector quantization Linear predictive coding Pattern matching	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents a means for achieving an 800-bits per second (b/s) speech communication capability by redigitizing the linear predictive coefficients present in the 2400-b/s data stream produced by the narrowband transmitter currently under development by the Department of Defense (DoD) and private industry. The 2400- to 800-b/s rate converter uses a vector quantization scheme to achieve most of the data reduction. This approach makes it possible to interconnect, without user intervention, 2400- and 800-b/s users (although both users would be effectively at the lower rate) via

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

rate converters located somewhere along the link. The intelligibility of the synthesized speech produced by this procedure compares well with other existing very-low-data-rate (VLDR) systems as determined by the Diagnostic Rhyme Test (DRT).

CONTENTS

2400- TO 800-b/s LPC RATE CONVERTER

INTRODUCTION

A high-quality 2400-bits per second (b/s) speech encoder using the principle of the linear predictive coder (LPC) has been developed by the Department of Defense (DoD) and private industry for narrowband voice communication over high-frequency channels, wirelines, and satellite links [1]. The 2400-b/s voice processor, known as LPC-10, is currently being specified for Federal Standard 1015 and Military Standard 188-113, and it will be the only narrowband device extensively deployed in the United States armed services and civilian government agencies. As well, the processor will be used by the armed forces of the North Atlantic Treaty Organization and other friendly countries.

Although LPC-10 is expected to adequately serve most future voice communication needs, some operational conditions need a very-low-data-rate (VLDR) capability (VLDR covers the transmission range from 600 to 1200 b/s). For example, speech may have to be transmitted through underwater channels, or in the presence of man-made interference, or at a low probability of intercept.

This report presents a means for achieving an 800-b/s capability by reencoding the 2400-b/s speech data. The rate converter was simulated in nonreal time on a PDP-11/45 and in real time on Navy-owned microprocessors. This report provides test results from each of these simulations.

The 800-b/s speech data preserves most of the intelligibility of the original 2400-b/s speech data. Intelligibility, as measured by the Diagnostic Rhyme Test (DRT), suffers only a 3.9 drop when going from 2400 to 800-b/s (for a single male speaker, called "RH," recorded in a quiet environment with a dynamic microphone).

The rate converter on the transmission side uses vector quantization of the reflection coefficients with a table containing 4096 reference vectors. For each set of reflection coefficients to be coded, a vector from the table is chosen which "best matches," in some predetermined sense, the coefficients: the index of that vector is transmitted. The rate converter on the receive side has the identical table from which the spectral information is retrieved.

The rate converter has the following advantages:

- It is a cost-effective means for achieving an 800-b/s capability.
- During overloaded or distrupted channel conditions, narrowband communication survivability could be increased by rate reduction to 800 b/s.
- Flexibility is introduced into the communication system by allowing interconnection between 2400- and 800-b/s users (although both users would be effectively at the lower rate).

Manuscript approved March 9, 1983.

PREVIOUS VLDR EFFORTS

Prior to presenting the 2400- to 800-b/s LPC rate converter, some previously reported VLDR efforts are reviewed in this section because some aspects are similar to the author's approach. Each effort is based on the narrowband vocoder model (i.e., channel vocoder, formant vocoder, or LPC) which chooses spectral information from a predetermined set. Such an approach has been termed "vector quantization," "pattern matching," or "block encoding."

VLDR Channel Vocoders

Phonetic Pattern Recognition Vocoder

The first application of pattern matching to speech coding was done by Dudley [2] in the late 1950s. An analog phonetic pattern-matching vocoder was developed that had 10 spectral patterns. Four consonants and six vowel sounds were used as the source for these patterns. Although intelligibility was quite limited, feasibility of pattern matching was demonstrated.

Spectral Pattern Matching Vocoder

More extensive testing was reported by Smith [3] in 1969. A channel vocoder was used to develop spectral patterns. Twelve different tables of spectral patterns were tested ranging from 497 patterns (328 voiced, 169 unvoiced) to 3996 patterns (2893 voiced, 1103 unvoiced). Speech compression from 450 to 600 b/s was achieved with reasonably good intelligibility reported.

Lincoln Laboratory 800-b/s Vocoder

A recent effort by Lincoln Laboratory uses the Spectral Envelope Estimation (SEE) Vocoder to perform the analysis portion of their real-time 800-b/s implementation [4]. Spectral information produced by SEE every 25 ms is treated as a vector (template). The "closest template" from a set of templates is found by means of an exhaustive search. The index of the "closest template" is transmitted to the receiver. The template set used is energy normalized. The gain is extracted and transmitted as a separate parameter.

The distance measure is a weighted difference between energy-normalized. log-spectral envelopes generated by the SEE vocoder:

$$d(s, t) = \sqrt{\frac{1}{42} \sum_{i=0}^{41} w(i) * [s(i) - t(i)]^2}, \tag{1}$$

where s and t are energy-normalized log-spectra spanning 0 to 3.8 kHz. The adaptive weighting function w is the maximum of the individual weighting functions:

$$w(i) = \max \{ws(i), wt(i)\}, \tag{2}$$

where ws and wt are individual weighting functions of the log-spectra s and t respectively. The individual weighting functions are determined from three inputs: model of the auditory masking, model of the resolution of hearing as a fixed function of frequency, and a weighting process which weights the high-energy regions more heavily than the low-energy regions.

The set of templates is updated in real time with new templates. If a new template's minimum distance to all the templates in the template set exceeds a threshold value, the new template is transmitted to the receiver. A longest-time-since-use algorithm decides which template in the set to discard. Two versions of this updating process have been tested: one is a continuous channel version and the other is a packetized channel version. The continuous channel system transmits new templates

during silence periods, and the packetized channel system transmits new templates by temporarily increasing the data rate. Changes in the speaker cause short-term degradation of the voice quality for the continuous channel system and short-term increase in the data rate for the packetized channel system. Both of these systems have a data rate slightly less than 800 b/s (packetized version may exceed 800 b/s during template updating). A three-male speaker DRT score of 83.9 was achieved for a version operating at 740 b/s without speaker adaptation [5].

VLDR Formant Vocoder

NRL 600-b/s Vocoder

Kang and Coulter [6] in 1976 used pattern matching on the output of a formant vocoder to encode the first three formants into seven bits per frame to generate synthesized speech at 600 b/s. A weighted difference measure was used which was defined as

$$D(i, j) = \sum_{m=1}^{3} [\{f(m, i) - f(m, j)\}\, w(m)]^2, \tag{3}$$

where

$$1 \leqslant i \leqslant 128,$$
$$1 \leqslant j \leqslant 128,$$
$$i \neq j,$$
$$w(m) = 4 - m,$$

and $f(m, i)$ is the mth formant frequency ($m = 1$, 2, and 3) of the ith pattern and $w(m)$ is the weighting factor for the mth formant frequency. The weighting factors emphasize the most important formant frequencies from a perceptual viewpoint. The third formant is least important because synthesized speech is intelligible in most cases with only the first two formants. Although the first and second formant are important, the first formant is weighted more heavily because its level is more constant and errors or fluctuations in its values are more obvious to the human ear. A single male speaker (speaker "CH") DRT score of 79.9 was achieved by this approach.

VLDR LPC Vocoders

NRL 1200-b/s Vocoder

The author of this report applied the pattern matching approach to linear predictive coefficients in 1975 to achieve a data rate of 1200 b/s [7]. The first two filter coefficients (ten in all) arising from the LPC analysis filter were quantized individually (5 and 4 bits for the first and second coefficients respectively). The remaining eight coefficients were treated as a pattern. The nearest neighbor pattern (using an Euclidean distance measure) was found from a stored table of 2048 patterns in memory; the index of this pattern was transmitted to the receiver which had an identical table of patterns to convert the index back into filter coefficients. The exclusion of the first two filter coefficients from the pattern-matching approach was a result of insufficient computer memory to include all ten filter parameters and of the need to adequately describe the first formant which is greatly influenced by the first two filter coefficients. A DRT score of 88.5 was achieved using a single male speaker (speaker "CH").

STI 800-b/s Vocoder

The application of vector quantization (pattern matching) to LPC filter coefficients with the likelihood-ratio distortion measure has been tested by Wong, Juang, and Gray [8]. The likelihood-ratio measure will be developed next.

Let the z transform of a frame of speech be denoted by $X(z)$, and the optimal 10th order LPC model for $X(z)$ be denoted by $\sqrt{\alpha_{10}}/A_{10}(z)$ where the residual energy term, α_{10}, results from inverse filtering $X(z)$ with $A_{10}(z)$. If $1/A(z)$ is any 10th order all pole filter, inverse filtering $X(z)$ with $A(z)$ results in a residual energy term α. Because α is minimized by $A_{10}(z)$, it follows that $\alpha_{10} \leqslant \alpha$. Residual energy, α, is given by

$$\alpha = \int_{-\pi}^{\pi} |X(e^{j\theta})|^2 \, |A(e^{j\theta})|^2 \, \frac{d\theta}{2\pi}. \tag{4}$$

For two unity gain model spectra, the likelihood-ratio measure is defined as

$$\begin{aligned} d(1/A_{10},\ 1/A) &= \int_{-\pi}^{\pi} |A(e^{j\theta})/A_{10}(e^{j\theta})|^2 - 1 \\ &= \frac{\alpha}{\alpha_{10}} - 1. \end{aligned} \tag{5}$$

Equation (5) shows that the measure is minimized when the residual energy is minimized.

A codebook of 1024 vectors was generated for the voiced case from a set of voiced training vectors. Another codebook of 1024 vectors was generated for the unvoiced case from a set of unvoiced training vectors. The codebooks were constructed such that the average spectral distortion from all the training vectors to their "best match" in the codebook was below a preset threshold [9]. A DRT score of 80.0 was achieved by using this approach for two male speakers (speakers "LL" and "CH").

TRW 800-b/s Vocoder

TRW achieved the 800-b/s data rate by block encoding orthogonalized reflection coefficients over four consecutive LPC frames [10]. Although the number of bits apportioned to each block was fixed, a dynamic bit allocation scheme was utilized to encode each parameter based on the voicing decision for the four frames within the block.

An intelligibility test score of 78.3 was achieved on the DRT for three male speakers ("LL," "CH," and "RH") recorded in a quiet background with a dynamic microphone. When unquantized parameters were replaced with quantized LPC-10 parameters prior to using their coding techniques (a 2400- to 800-b/s rate converter simulation), it was found that the DRT score for the same three male speakers dropped to 71.9.

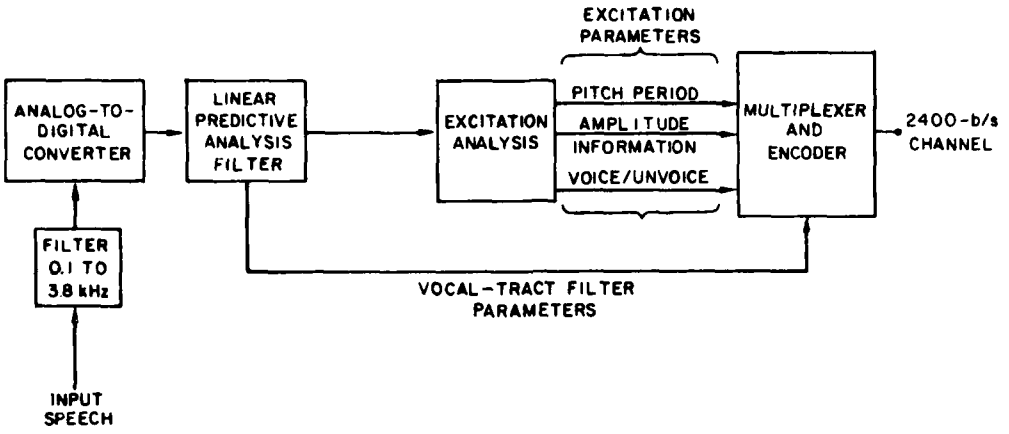
OVERVIEW OF 2400- TO 800-b/s RATE CONVERTER

Block Diagram of LPC-10

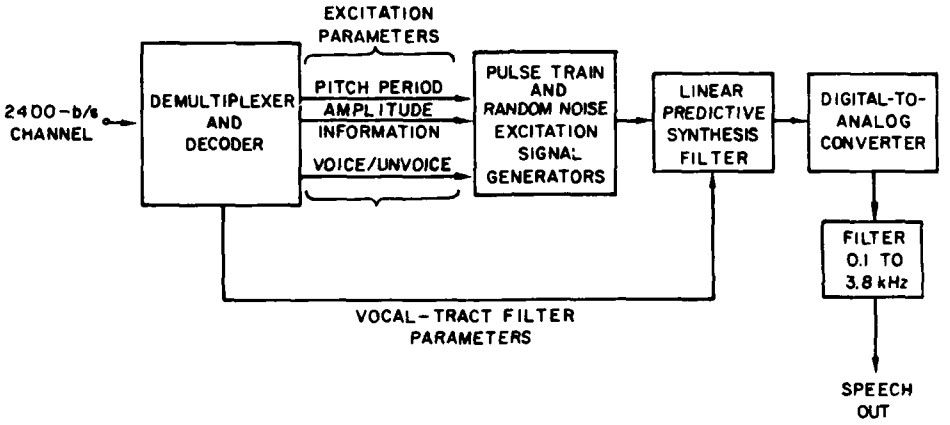
LPC-10 encodes the speech waveform into two sets of parameters. One set consists of vocal-tract-filter parameters (spectral coefficients or reflection coefficients), estimated by the least-squares method, which describes the signal transformation (resonant frequency) characteristics of the vocal tract. The other set describes the excitation waveform consisting of amplitude, pitch period, and voiced/unvoiced decision (buzz/hiss selection). Both sets of parameters are derived once every 22.5 milliseconds (ms) and quantized to 54 bits. At the receive end, reflection coefficients are fed into the synthesis filter which is excited by the reconstructed excitation signal as shown in Fig. 1.

Features of the 2400-b/s voice processor that are considered desirable to maintain with the rate converter in the link are:

- operate on casual conversation,
- suitable for multispeaker environment without individualized "training" for particular speakers,



(a) Transmitter



(b) Receiver

Fig. 1 — LPC-10 voice processor

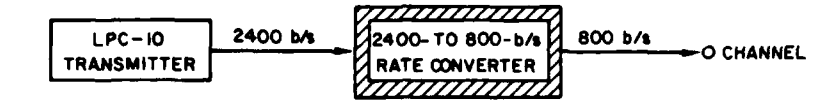
- function with processing delay of not more than half of a second (delays of more than half of a second make it difficult to maintain two-way conversations).

Block Diagram of Rate Converter

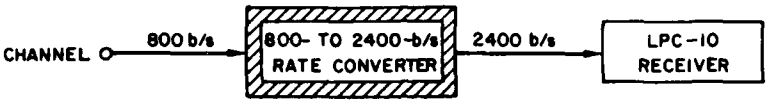
The rate converter requantizes the already quantized parameters coming out of the narrowband transmitter as shown in Fig. 2. The rate converter could be pulled into the 2400-b/s transmitter to quantize the "raw" (unquantized) signals directly. This would result in a higher quality 800-b/s synthesized speech, but this approach would not take advantage of the planned extensive deployment of the 2400-b/s processor.

Bit Allocation of Encoded LPC-10 Parameters

The 800-b/s capability was developed in accordance with two constraints: first, the output bit stream of the 2400-b/s transmitter would be the starting point for further data reduction; and second,



(a) Transmit end



(b) Receive end

Fig. 2 — Rate converter integrated with LPC-10

the algorithm complexity should be minimized to enhance the possibility for hardware implementation. Pertinent details of the 2400-b/s LPC are listed in Table 1.

Bit Allocation of Encoded 800-b/s Parameters

Similar to some previous VLDR vocoders [3-8], the present rate converter transmits all parameters once per frame, except the pitch period. Since the rate of change of the pitch period in normal conversation is much lower than either amplitude or spectral information, it may be transmitted for every three LPC frames without introducing noticeable unnaturalness. Table 2 lists the allocation bits for the VLDR parameters.

Table 1 — LPC-10 Design Parameters

General Information			
Speech sampling rate (kHz)			8
Frame rate (Hz)			44.444
Frame size (speech samples)			180
Encoded Data (bits/frame)			
Sync bit			1
Excitation parameters			
Amplitude			5
Pitch period			6
Voicing decision			1
Synthesis filter coefficients		(if voiced)	(if unvoiced)
Coefficient	#1	5	5
	2	5	5
	3	5	5
	4	5	5
	5	4	0
	6	4	0
	7	4	0
	8	4	0
	9	3	0
	10	2	0
Error-protection codes		0	20
Unused bit		0	1
		Total	54 bits/frame

Table 2 — Frame Rate and Bit Allocation of the 2400- and 800-b/s Data Rates

	Data Rate	
	2400 b/s	800 b/s
Timing Information		
Frame Rate	44.44 frames/s	14.81 frames/s
Frame Time	22.5 ms	67.5 ms
Encoded Data (bits/frame)		
Synchronization	1	1
Excitation Parameters		
Amplitude	5	4 + 4 + 4
Pitch	6	5
Voicing	1	[a]
Filter Parameters	41	12 + 12 + 12
Total	54	54

[a]Voicing decision included with filter parameter index

Excitation Encoding

From an intelligibility standpoint, voicing and amplitude information are more important than pitch (good pitch quality aids primarily in making more natural sounding speech). Thus, voicing and amplitude parameters are updated at the same frequency in both the 800- and 2400-b/s data rates. The voicing decision and filter information are coded together and will be discussed in the next section. The 5-bit amplitude information at the 2400-b/s rate is mapped to 4 bits for the 800-b/s rate as shown in Table 3.

In the 800-b/s rate, pitch information is updated once for every three 2400-b/s frames. Also, six bit pitch information at the 2400-b/s rate is mapped to 5 bits for the 800-b/s rate as shown in Table 4.

FILTER PARAMETER ENCODING

Coarser quantization of both the excitation and filter parameters to achieve the 800-b/s rate results in degraded synthesized speech. The effect of the coarser quantized excitation parameters, however, is relatively minor. As discussed in the preceding section, the 800-b/s vocoder encodes the amplitude parameter at 4 bits, 1 bit less than the 2400-b/s LPC. Reduction of 1 bit in the amplitude resolution (i.e., 1.5 dB to 3.0 dB resolution) is not easily detectable in the synthesized speech by casual listening. Likewise, the 800-b/s vocoder transmits the pitch period at 5 bits, 1 bit less than the 2400-b/s LPC. Although the rate converter updates the pitch parameter at one-third the rate of LPC-10 (i.e., at a rate of 14.81 times per second), the slower pitch update will rarely introduce unnatural intonations in the synthesized speech during normal conversational speech. In summary, data reduction for the excitation parameters does not introduce significant speech degradation.

On the other hand, the effect of data rate reduction on the filter coefficients is significant. As noted from Table 1, LPC-10 transmits one out of 2^{41} possible spectral sets when speech is voiced. These 2.2 trillion spectral sets are reduced to 4000 or less in the 800-b/s vocoder. This 500,000,000-to-one reduction of spectral sets is a significant cause of speech degradation, particularly for a multispeaker environment with casual conversational speech.

Table 3 — Amplitude Coding Table

Preemphasized Speech RMS Value	2400-b/s Amplitude Code	800-b/s Amplitude Code	Regenerated 2400-b/s Amplitude Code
468 or more	31	15	30
392-467	30	15	30
328-391	29	14	28
275-327	28	14	28
230-274	27	13	26
192-229	26	13	26
164-191	25	12	24
135-163	24	12	24
113-134	23	11	22
94-112	22	11	22
79-93	21	10	20
66-78	20	10	20
55-65	19	9	18
46-54	18	9	18
39-45	17	8	16
32-38	16	8	16
27-31	15	7	14
23-26	14	7	14
19-22	13	6	12
16-18	12	6	12
13-15	11	5	10
11-12	10	5	10
9-10	9	4	8
8	8	4	8
7	7	3	6
6	6	3	6
5	5	2	4
4	4	2	4
3	3	1	2
2	2	1	2
1	1	0	0
0	0	0	0

Table 4 — Pitch Coding

Pitch Period	2400-b/s Pitch Code	800-b/s Pitch Code	Regenerated 2400-b/s Pitch Code
Unvoiced	0		0
20	19		30
21	11	0	30
22	27	0	30
23	25	0	30
24	29	0	30
25	21	0	30
26	23	0	30
27	22	0	30
28	30	0	30
29	14	1	15
30	15	1	15
31	7	2	39
32	39	2	39
33	38	3	46
34	46	3	46
35	42	4	43
36	43	4	43
37	41	5	45
38	45	5	45
39	37	6	53
40	53	6	53
42	49	7	49
44	51	8	51
46	50	9	50
48	54	10	54
50	52	11	52
52	60	12	60
54	56	13	56
56	58	14	58
58	26	15	26
60	90	16	90
62	88	17	88
64	92	18	92
66	84	19	84
68	86	20	86
70	82	21	82
72	83	22	83
74	81	23	81
76	85	24	85
78	69	25	69
80	77	26	77
84	73	27	73
88	75	28	75
92	74	29	74
96	78	30	70
100	70	30	70
104	71	31	67
108	67	31	67
112	99	31	67
116	97	31	67
120	113	31	67
124	112	31	67
128	114	31	67
132	98	31	67
136	106	31	67
140	104	31	67
144	103	31	67
148	100	31	67
152	101	31	67
156	76	31	67
Voicing Transition	127		

In spite of the severe reduction of spectral information in going from 2400 to 800 b/s, the synthesized speech from the 800-b/s speech data still retains most of the intelligibility of the original 2400-b/s speech data. Some reasons for this are:

- The encoding table is constructed only from human voice sounds. LPC-10 can reproduce nonspeech-like signals: noise, animal sounds, or any other acoustic signals which have three or four resonant frequencies within the passband. The vector quantization table, however, does not contain these spectra.
- Additional bit-saving by vector quantization stems from a many-to-one mapping property inherent in vector quantizing filter coefficient sets. Many different vectors produce a similar spectrum which may be mapped into one vector. This is particularly true when lower indexed reflection coefficients have large magnitudes (i.e., overall spectrum is essentially unaffected by higher indexed reflection coefficients).

Distance Measures

There have been numerous distance measures reported in literature which can be implemented in a 800-b/s vocoder utilizing vector quantization. Some of these distance measures are:

1. Itakura-Saito [11]
2. Likelihood ratio (gain-normalized Itakura-Saito) [8,12]
3. Viswanathan [13]
4. Cepstral [14]
5. Log-area ratio [13,15]

The most meaningful distance measures would be those that indicate the perceived difference. None of the above distance measures is based on known or measurable properties of auditory perception (i.e., the masking effect, the logarithmic resolution of frequency, and the like). In fact, many of these measures produce similar results, hence the computationally most efficient log-area ratio is much preferred [13].

This report tests two well-used distance measures in terms of DRT (which has not been done before): likelihood ratio and log-area ratio. The intelligibility scores of the 800-b/s vocoder based on these two distance measures are given in the next section dealing with experimentation. The definition of the likelihood-ratio measure is given in Eq. (5), and the definition of the log-area-ratio distance is

$$D(i, j) = \sum_{m=1}^{10} (y(m, i) - y(m, j))^2 \tag{6}$$

where

$$y(i) = \log \frac{1 + k(i)}{1 - k(i)} \text{ for } k(i) \geq 0,$$

$$= -\log \frac{1 - k(i)}{1 + k(i)} \text{ for } k(i) < 0.$$

The $k(i)$ is the ith reflection coefficient generated by the LPC analyzer.

Table Generation

In the past, two different table-generation techniques have been employed in 800-b/s vocoders. STI's 800-b/s LPC vocoder [8] used the table generated by the K-mean algorithm which is well known in the pattern recognition field [16]. On the other hand, Lincoln Laboratory's real-time 800-b/s channel vocoder [4] used a nonclustering approach. In this approach, the table is so generated that the distance between any two different vectors in the table is greater than a fixed value (M); namely,

$$d(i, j) > M \text{ for } i \neq j \tag{7}$$

where $d(i, j)$ is the distance between the ith and jth vector in the table. Ideally, the choice of M is made such that there is a just-noticeable perceptual difference between the tones generated by different vectors in the table.

The task of comparing these two table-generation techniques is immense. It is significant to note, however, that both the STI and Lincoln Laboratory approaches produce high-quality speech by using two different table-generation techniques.

The approach used by Lincoln Laboratory, however, has a decided advantage from an operational viewpoint by allowing real-time adaptation of the table to the speaker. With real-time updating of the table, the speech quality at 800 b/s is nearly as good as that of 2400 b/s. Because of this advantage, the table-generation method by Lincoln Laboratory was selected for further experimentation.

Speech Source for Table Generation

Intelligibility of the 800-b/s simulation was found to be sensitive to the kind of speech input used to generate the table of vectors. Better performance was obtained with tables generated from many speakers. For this reason, a recording on analog tape was made at NRL of 54 males and 12 females each speaking five different phonetically balanced sentences. From this master tape, another tape was produced with the same speakers each speaking two sentences. These speech data were run through an LPC-10 analyzer to generate a set of voiced training vectors and a set of unvoiced training vectors.

Total Number of Vectors

As indicated by Table 2, 12 bits or 4096 vectors are allowed for LPC coefficients, for both voiced and unvoiced speech. Since unvoiced speech does not require accurate spectral representation, the number of voiced vectors can be greater than the number of unvoiced vectors. This trade-off will be made in the next section.

EXPERIMENTAL RESULTS FROM NONREAL-TIME SIMULATION

The rate converter was simulated in nonreal time with a FORTRAN program running on a PDP-11/45 (details are given in Appendix A).

Distortion Measure: Likelihood Ratio vs Log Area Ratio

Although there are many distortion measures that could be used in the rate converter algorithm, only the well-known likelihood ratio and log-area ratio were tested. The likelihood ratio has some desirable theoretical properties, such as: it is a gain-normalized-spectral distortion measure and minimizing this measure is equivalent to minimizing the residual energy. The log-area ratio has some advantageous computational considerations when compared to the likelihood ratio, such as: half the number of multiplications for determining "closest match" and half the amount of memory needed for the voiced and unvoiced tables. Test results of Table 5 compare the likelihood ratio against the log-area ratio. Because of the importance of choosing a good measure, a three-male speaker (speakers "LL,"

Table 5 — Comparison of the Likelihood Ratio and Log-Area Ratio Distortion Measures

Algorithm or Distortion Measure	Data Rate (b/s)	Condition of RCs Prior to Pattern Matching	No. of Vectors		Speaker(s)	DRT Score
			Voiced Table	Unvoiced Table		
LPC-10	2400	—	—	—	LL, CH, RH	88.4
Likelihood Ratio	800	Unquantized	3072	1024	LL, CH, RH	82.8
Log-Area Ratio	800	Unquantized	3072	1024	LL, CH, RH	83.2

"CH," and "RH") DRT was run for this comparison. DRT scores for both measures are within the standard error of each other. From an intelligibility standpoint, there is no clear choice for choosing one of these two measures. Hence, only the less computationally demanding log-area ratio was considered in subsequent experiments.

Six initial consonant attributes are tested by the DRT. Figure 3 shows how the distortion measures compared by attribute. As can be seen in Fig. 3, the distinctions between the two measures are not significant. The greatest disparity between the 2400- and 800-b/s rates occurs with the graveness attribute which tests in the present condition initial consonants with relatively low position of second and third formants and in the absent condition initial consonants with relatively high position of second and third formants. Because the greater part of the data reduction in going from 2400 to 800 b/s takes place with the filter parameters, it would be expected that the greatest degradation would occur in those attributes which are most influenced by the formant structure.

Table Size: Voiced vs Unvoiced

By allowing the filter parameter index to also determine the voicing decision, a larger voiced table than unvoiced table is made possible. This is desirable because a coarse spectral representation is adequate for the unvoiced state. For the 800-b/s case, 12 bits determine the filter parameters which allows for a table of 4096 vectors. Just how large the voiced table should be was determined through informal listening tests and DRT scores. Table 6 lists scores for two different voiced/unvoiced table size combinations. Only a single speaker, called "RH," was run here. Speaker "RH" was chosen because he generally scored somewhere between "LL" and "CH" (middle of the road). Based on these scores a voiced table size of 3840 (index 0 through 3839) and an unvoiced table size of 256 (index 3840 through 4095) was chosen.

Reflection Coefficients: Unquantized vs Quantized

The scores presented so far have been for the rate converter operating on unquantized reflection coefficients (rate converter pulled inside the LPC-10 transmitter). A loss of intelligibility does occur when the rate converter is pulled outside the LPC-10 where it requantizes the already quantized reflection coefficients in the the 2400-b/s data stream. For speaker "RH," the DRT score drops 1.4 (as shown in Table 7).

EXPERIMENTAL RESULTS FROM REAL-TIME SIMULATION

A real-time simulation of the rate converter was performed on Navy-owned microprocessors (details are given in Appendix B). The number of vectors in the voiced table of the real-time simulation was limited to 1024 vectors because of speed and memory limitations of the hardware. The intelligibility scores in this section are lower because of these hardware limitations.

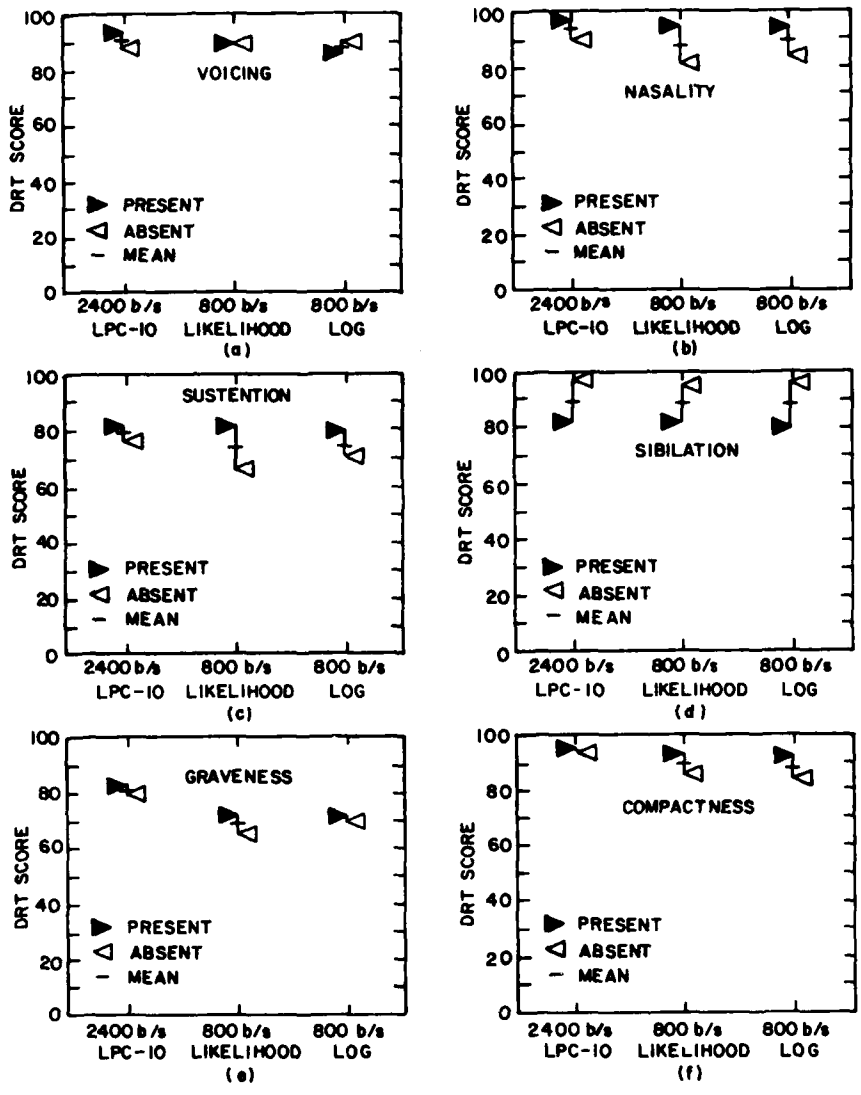


Fig. 3 — Comparison of distortion measures by consonant attributes tested by the DRT

Table 6 — Results Obtained from Two Different Voiced/Unvoiced Table Configurations

Algorithm or Distortion Measure	Data Rate (b/s)	Condition of RCs Prior to Pattern Matching	No. of Vectors		Speaker(s)	DRT Score
			Voiced Table	Unvoiced Table		
LPC-10	2400	—	—	—	RH	87.1
Log-Area Ratio	800	Unquantized	3840	256	RH	84.6
Log-Area Ratio	800	Unquantized	3072	1024	RH	83.2

Table 7 — DRT Scores for Unquantized and Quantized Reflection Coefficients Prior to Pattern Matching

Algorithm or Distortion Measure	Data Rate (b/s)	Condition of RCs Prior to Pattern Matching	No. of Vectors		Speaker(s)	DRT Score
			Voiced Table	Unvoiced Table		
LPC-10	2400	—	—	—	RH	87.1
Log-Area Ratio	800	Unquantized	3840	256	RH	84.6
Log-Area Ratio	800	Quantized	3840	256	RH	83.2

Table Formation: Smaller vs Larger Distance (M)

The choice for a value to give to the parameter M of Eq. (7) was arrived at initially through informal listening tests. Perceptual differences for small changes in M where found to be small. Increasing M results in vectors being more "spread out" and requires additional speech input during table generation. Table 8 lists some DRT results obtained for three different choices for the parameter M.

Table 8 — Comparison of DRT Scores for Various Values of the Input Parameter M

Distortion Measure	Data Rate (b/s)	Speech Source Used During Table Generation			DRT[a] Score
		No. of Male/Female Speakers	No. of Sentences Spoken by Each Speaker	No. of Sentences Used	
Log-Area Ratio	800	54M/11F	5	325	79.5
Log-Area Ratio	800	52M/11F	2	126	80.7
Log-Area Ratio	800	35M	2	70	80.7

[a]Speakers "LL," "CH," and "RH."

Vector Size: Ten vs Eight Reflection Coefficients

For the voiced case, fixing the last two coefficients and letting a vector consists of $y(1)$ through $y(8)$ results in a "closer match" of the first eight coefficients. On the other hand, $y(9)$ and $y(10)$ are quantized roughly. DRT runs were made comparing vector sizes of $y(1)$ through $y(8)$ against $y(1)$ through $y(10)$. The scores for the two approaches were insignificantly different (78.0 and 78.3 for a vector size of eight and ten respectively).

Results Obtained When Current User's Voice Was Also Used to Generate Table

Intelligibility of the 800-b/s speech is significantly improved if the table of vectors was generated from the person currently using the system. It is possible to adapt the table of vectors to the individual currently speaking (i.e., sending new vectors to the receive side during silence periods). Improvement that could be expected by this real-time adaptation is presented next. Table 9 lists several DRT scores obtained from a single male speaker called "LL." Speaker "LL" was used to generate two of the tables (cases 2 and 3). DRT scores for these two cases are significantly improved. These scores represent the upperbound in performance of the system, for speaker "LL," if the vectors in the table were updated in real time. Note that the voiced table size can be dropped from 1024 to 256 vectors while still maintaining performance better than the first case which has a table from which no vectors originated from speaker "LL."

Table 9 — Intelligibility Improvement Found when DRT Speaker Was Also Used to Generate Table

Case	Distortion Measure	Data Rate (b/s)	No. of Vectors in Table	Speech Source Used for Generation of Table	Speaker "LL"[a] DRT Score
1	Log-Area Ratio	800	1024	54M/11F each speaking 2 sentences	75.8
2	Log-Area Ratio	800	1024	Speaker "LL" used to generate table	79.2
3	Log-Area Ratio	800	256	Speaker "LL" used to generate table	77.5

[a]One male speaker called "LL"

Table Generation: With vs Without Background Noise

If the rate converter is to be operated in a noisy environment, the question arises as to whether the table should be generated with the same background noise present. Although the results presented next do not definitively answer this question, they point in the direction of comparable intelligibility being achieved by having the table formed from "clean" speech. Scores achieved in Table 10 were obtained, in all cases, by running the DRT with three male speakers using a M87 microphone with RH-53 helicopter background noise present. The DRT tape was not used during table generation. In each of the three cases, table construction differed. In case 1, the table was formed with the helicopter background noise present. In case 2, the same speech source was used to generate the table but without the helicopter noise. Cases 1 and 2 have the same parameter M value, but it took considerably less speech input to generate the table in case 2. In case 3, the M value was increased to achieve the same amount of speech input for generating the table as was necessary in case 1. In case 3, the DRT score for the table formed from "clean speech" did better than the table generated with "dirty" speech.

Table 10 — DRT Results with Helicopter Background Noise Present

Case	Distortion Measure	Data Rate (b/s)	Speech Source Used During Table Generation				DRT[a] Score
			Helicopter Noise Present	No. of Male Speakers	No. of Sentences Spoken by Each Speaker	Total No. of Sentences Used	
1	Log-Area Ratio	800	yes	51	2	102	57.8
2	Log-Area Ratio	800	no	20	2	40	53.4
3	Log-Area Ratio	800	no	51	2	102	58.7

[a]Three male speakers with helicopter background noise present.

CONCLUSIONS

This report describes a practical approach to achieving an 800-b/s voice communication capability. The 800-b/s data rate is achieved by requantizing the 2400-b/s speech data stream of LPC-10. Some of the benefits of such an approach to achieving a VLDR capability are:

- the rate converter takes advantage of the planned extensive deployment of LPC-10;
- communication between 800 and 2400-b/s users is made possible;

- network connection during overloaded channel conditions can be made by converting to the 800-b/s data rate.

ACKNOWLEDGMENTS

The author expresses his appreciation to Mr. George S. Kang for his many thoughtful suggestions and encouragements. Also, the author gratefully acknowledges the support of Mr. Robert Martin of the Naval Electronic Systems Command.

REFERENCES

1. G.S. Kang, L.J. Fransen, and E.L. Kline, "Multirate Processor (MRP) for Digital Voice Communications," Appendix C of NRL Report 8295, Mar. 21, 1979.

2. H. Dudley, "Phonetic Pattern Recognition Vocoder for Narrow-Band Speech Transmission," J. Acoustic Soc. Am. **30**, 733-739, Aug. 1958.

3. C.P. Smith, "Perception of Vocoder Speech Processed by Pattern Matching," J. Acoustic Soc. Am. **46**, 1562 (1969).

4. D.B. Paul and P.E. Blankenship, "Two Distance Measure-Based Vocoder Quantization Algorithms for Very-Low Data Rate Applications: Frame-Fill and Spectral Vector Quantization," International Computer Conference, Philadelphia, June 1982.

5. D.B. Paul, "Frame-Fill and Vector Quantization at 500-800 BPS," talk given at Digital Voice Processor Consortium Workshop, Mitre Corp., McLean, Virginia, Oct. 1982.

6. G.S. Kang and D.C. Coulter, "600 bps Voice Digitizer," pp. 91-94 in Conference Record of 1976 IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, Pa., Apr. 12-14.

7. L.J. Fransen, "Application of Pattern Matching to Linear Predictive Coding of Speech at 1200 bps," NRL Report 7931, Oct. 1975.

8. D.Y. Wong, B.H. Juang, and A.H. Gray, Jr., "An 800 bit/s Vector Quantization LPC Vocoder," IEEE Trans. on Acoustics, Speech, and Signal Processing **ASSP-30** (5), Oct. 1982.

9. B.H. Juang, D.Y. Wong, and A.H. Gray, Jr., "Distortion Performance of Vector Quantization for LPC Voice Coding," IEEE Trans. Acoustics, Speech, and Signal Processing **ASSP-30** (2), April 1982.

10. T.E. Carter, D.M. Dlugos, and D.C. leDoux, "An 800 b/s Real-Time Voice Coding System Based on Efficient Encoding Techniques," in Proc. 1982 IEEE International Conference on Acoustics, Speech, and Signal Processing.

11. F. Itakura and S. Saito, "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies," Electron. Commun. Japan **53-A**, 36-43 (1970); also in *Speech Synthesis*, J.L. Flanagan and L.R. Rabiner, Ed. Stroudsburg, Pa. Dowden, Hutchinson, and Ross, 1973, p. 293-304.

12. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoustics, Speech, and Signal Processing, **ASSP-23**, 145-150, Feb. 1975.

13. R. Viswanathan, W. Russell, and J. Makhoul, "Objective Speech Quality Evaluation of Narrowband LPC Vocoders," Record of 1978 IEEE International Conference on Acoustics, Speech, and Signal Processing, 78CH1285-6 ASSP, pp. 591-594.

14. T.P. Barnwell, A.M. Bush, R.M. Mersereau, and R.W. Schafer, "Speech Quality Measurement," 1977, Final Report Prepared for Rome Air Development Center, RADC-TR-78-122.

15. R. Viswanathan and J. Makhoul, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," IEEE Trans. Acoustics, Speech, and Signal Processing, June 75, pp. 309-321.

16. M.R. Anderberg, *Cluster Analysis for Applications*, New York: Academic Press, 1973.

Appendix A
FORTRAN SIMULATION OF RATE CONVERTER

The rate converter was simulated in nonreal time in FORTRAN on a PDP 11/45. A block diagram of the FORTRAN simulation is shown in Fig. A1. Frame rate and bit allocation of the 2400- and 800-b/s rates are the same as that given in Table 2. FORTRAN versions differed in how the reflection coefficients were quantized. These simulations helped determine:

- distortion measure to use
- number of vectors to be included in voiced and unvoiced tables
- improvement in intelligibility gained when vector-matching-unquantized-reflection coefficients (available if the rate converter is pulled inside the LPC-10 transmitter) rather than the quantized-reflection coefficients (present in the 2400-b/s LPC-10 data stream).

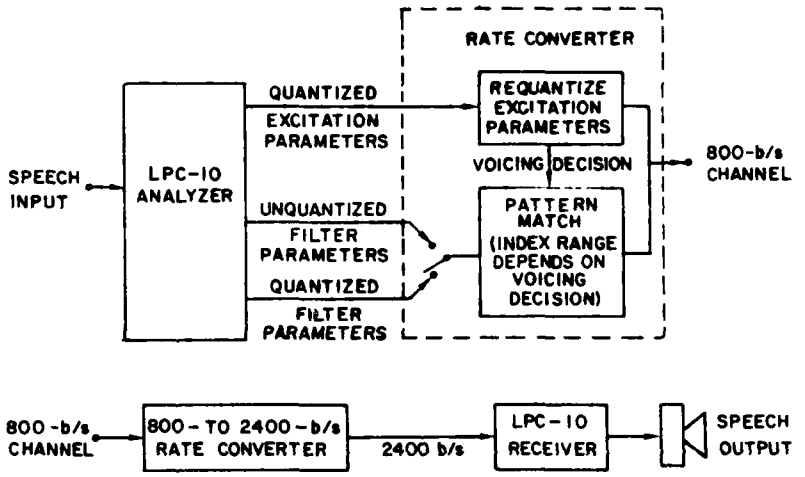


Fig. A1 — Block diagram of FORTRAN rate converter simulation

The analog tape of 54 males and 12 females (each speaker speaking two sentences) was used as the input to the analyzer of LPC-10 for generating the voiced and unvoiced training sets. Although the different algorithms tested required different voiced and unvoiced tables, in each case the tables were constructed such that the requirements of Eq. (7) were met. Also, the threshold value M was adjusted so that nearly all of the vectors in the appropriate training set were evaluated prior to filling the table.

Appendix B
REAL-TIME SIMULATION OF RATE CONVERTER

The 800-b/s rate converter was simulated in real time by using two Navy-owned microprocessors. These microprocessors were originally built to demonstrate the LPC-10 algorithm operating at 2400 b/s. The microprocessor configuration for the 800-b/s simulation is shown in Fig. B1. The right-hand microprocessor is operating in a full-duplex mode with the rate reduction being done in the left-hand unit.

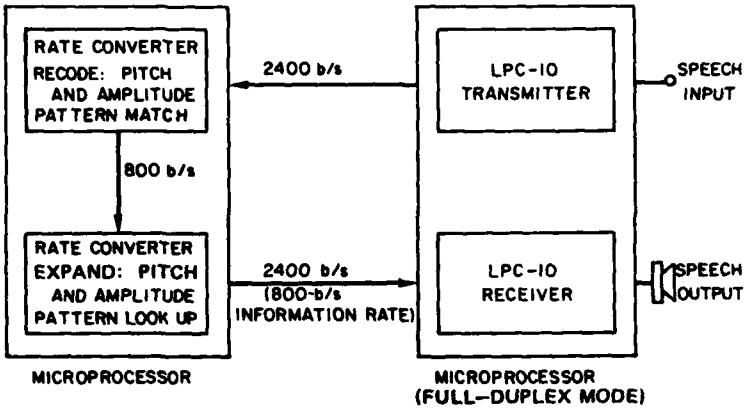


Fig. B1 — Hardware configuration for 2400 to 800-b/s rate converter

A maximum table size of 1024 vectors is realizable with this microprocessor because of memory size and hardware speed limitations. These limitations resulted in the real-time simulation algorithm being somewhat different from the FORTRAN simulation description already given where these considerations did not exist. For the real-time simulation, the 800-b/s frame is 36 bits and it is updated at 22.22 times per second. Pitch and synchronization parameters are updated once per 800-b/s frame. All other parameters are updated twice per 800-b/s frame. Frame bit allocation is shown in Table B1. The "closest match" for the voiced case was found by an exhaustive search of 1024 vectors.

Table B1 — Frame Bit Allocation for the 2400- and 800-b/s Real-Time Simulation

	Data Rate	
	2400 b/s	800 b/s
Timing Information		
Frame Rate	44.44 frames/s	22.22 frames/s
Frame Time	22.5 ms	45.0 ms
Encoded Data (bits/frame)		
Synchronization	1	1
Excitation Parameters		
Amplitude	5	4 + 4
Pitch	6	5
Voicing	1	2
Filter Parameters	41	10 + 10
Total	54	36